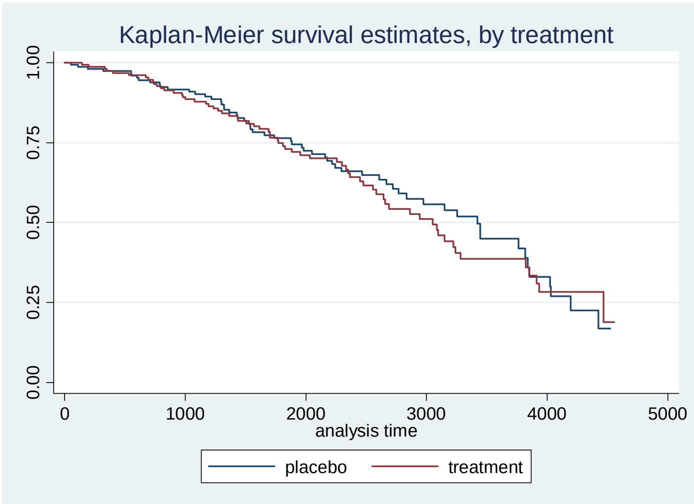
IGP 304 (Spring 2006) Homework 5 Keys:

1. In the data, there are three variables. For each subject, `treatment` indicates treatment (an experimental drug) or placebo, `obstime` is the days observed after treatment, and `status` is an indicator variable for whether the time observation was censored or not. Graph the survival curves for those patients that received the placebo, and those that received the experimental drug. Did the drug improve the survival? [Note: In Stata, use `stset obstime, failure(status)` to declare the outcome variables. You also may need to generate an indicator variable for treatment.]

The survival curves for the two groups are below. [In Stata, after you import the data, use `stset obstime, failure(status)` to declare the outcome variables and then `sts graph, by(treatment)` to draw the survival curves.]



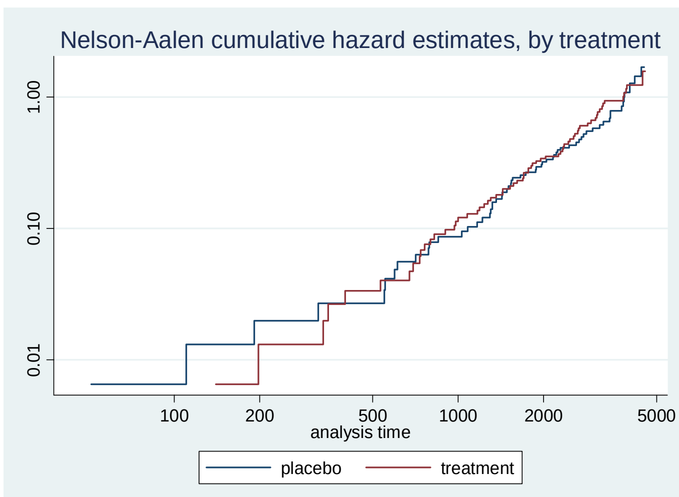

To test if the two survival curves are different, we can use either the Mantel-Cox method or Cox's proportional hazard regression; they give similar results. The results of the Mantel-Cox method is

```
. ** Define an indicator variable for treatment
. gen aa = 0
. replace aa = 1 if treatment == "treatment"
. stmc aa

         failure _d:  status
   analysis time _t:  obstime

Mantel-Cox comparisons

Mantel-Haenszel estimates of the rate ratio
   comparing aa==1 vs. aa==0
   controlling for time (by clicks)

Overall Mantel-Haenszel estimate, controlling for time
```

```
----------------------------------------------------------
     RR        chi2         P>chi2       [95% Conf. Interval]
----------------------------------------------------------
   1.075       0.16         0.6884          0.756      1.529
----------------------------------------------------------
```

The hazard ratio of the treatment group over the placebo group is estimated to be 1.075 with 95% CI [0.756, 1.529]. The test for this hazard ratio being one has p-value 0.69, and thus there is no evidence to support that the two groups have different hazard functions.

In either the Mantel-Cox method or the proportional hazard regression, we assume the hazard functions for the two groups have a constant ratio. We need to check the validity of this assumption. Since constant hazard ratio is equivalent to constant difference between log-cumulative hazards, we examine cumulative hazards plotted in log scales (see below). The two curves are mainly parallel, indicating that the assumption is met, except maybe for the very early time, where the variations of the estimates are high. [Stata command `sts graph, by(treatment) na xscale(log) yscale(log) xlabel(100 200 500 1000 2000 5000) ylabel(.01 .1 1)`, or `sts graph, by(treatment) xscale(log) yscale(log) h cih`]




2. A common symptom of otitis media in young children is the prolonged presence of fluid in the middle ear, known as middle-ear effusion. The presence of fluid may result in temporary hearing loss and interfere with normal learning skills in the first two years of life. One hypothesis is that babies who are breast-fed for at least 1 month build up some immunity against the effects of the disease and have less prolonged effusion than do bottle-fed babies. A small study of 24 pairs of babies is set up, where the babies are matched on a one-to-one basis according to age, sex, socioeconomic status, and type of medications taken. One member of the matched pair is a breast-fed baby, and the other member is a bottle–fed baby. The outcome variable is the duration of middle-ear effusion (in days) after the first episode of otitis media. The results are here.
   - Why might a nonparametric test be useful here? Which nonparametric test should be used here?

This study has a paired set-up, for which a paired t-test could be a choice of analysis. However, the distribution of the 24 differences is much skewed, with some vary large numbers; the same also is true for the ratios. Thus, nonparametric approaches will be useful here since they are insensitive to extreme values. Wilcoxon's signed rank test is appropriate for the data.

- Test the hypothesis that the duration of effusion is different among breast-fed babies than among bottle-fed babies using a nonparametric test.

The differences and ranks of the absolute values of the differences (excluding zero) are

| | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -2 | 24 | 4 | 158 | -1 | 5 | 165 | 0 | 18 | 59 | 169 | 17 | -13 | 1 | 9 | -2 | 1 | 12 | 12 | 2 | 1 | -3 | -6 | -5 |
| 6 | 19 | 9 | 21 | 2.5 | 10.5 | 22 | | 18 | 20 | 23 | 17 | 16 | 2.5 | 13 | 6 | 2.5 | 14.5 | 14.5 | 6 | 2.5 | 8 | 12 | 10.5 |

Adding up the ranks for positive and negative differences separately, we get $T_+ = 215$ and $T_- = 61$. Thus, the signed rank test statistic is $T = \min\{T_+, T_-\} = 61$. If we look up EMS Table A7 at row $N = 23$, we will find the p-values for $T = 62$ and 55 are 0.02 and 0.01, respectively. Hence, the p-value for $T = 61$ should be near 0.02, and there is a statistically significant evidence for the difference in duration of effusion between breast-fed than bottle-fed babies. [In Stata, you may use command `signrank breast = bottle`. This command will not exclude zero when ranking the differences, resulting in different values for the sums of ranks. But it also looks up a different reference distribution. Thus, the p-value should be the same as in our approach.]

3. A dietary survey was mailed to 537 students on two separate occasions, several months apart, including questions on many food items. Consider the red meat consumption responses listed below:

| | Survey 2 | | |
|---|---|---|---|
| Survey 1 | < 1 serving/wk | > 1 serving/wk | Total |
| < 1 serving/wk | 136 | 92 | 228 |
| > 1 serving/wk | 69 | 240 | 309 |
| Total | 205 | 332 | 537 |

Examine the reproducibility of the survey using the Kappa statistic. Interpret your outcome.

The expected numbers of concordant answers under independence are $205 \times 228 / 537 = 87.04$ for "<1 serving/wk" and $332 \times 309 / 537 = 191.04$ for ">1 serving/wk", respectively. The kappa statistic is $[(136 + 240) - (87.04 + 191.04)] / [537 - (87.04 + 191.04)] = 0.3782$. Since the kappa statistic is not high, the level of reproducibility of the survey was not high. But this doesn't mean there was no reproducibility. The responses of the two surveys were significantly associated, but the level of dependence was not strong. [When sample size is big enough, two weakly associated variables will show significant association.]